GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: December 29, 2004, 14:44:37 ; Search time 157 Seconds
(without alignments)
59.407 Million cell updates/sec

Title: US-09-998-042-1
Perfect score: 147
Sequence: 1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	147	100.0	26	4	AAG62388	Aag62388 Alternati
2	147	100.0	26	4	AAB48914	Aab48914 Human ace
3	147	100.0	26	4	AAB50032	Aab50032 Acetylcho
4	147	100.0	26	5	AAU98024	Aau98024 Human rea
5	147	100.0	31	2	AAW74588	Aaw74588 Amino aci
6	147	100.0	31	2	AAW68146	Aaw68146 Human ACh
7	147	100.0	37	4	AAB50035	Aab50035 Acetylcho
8	147	100.0	53	4	AAB50036	Aab50036 Acetylcho
9	147	100.0	53	5	ABG31331	Abg31331 GFP-fused
10	147	100.0	600	2	AAW48797	Aaw48797 Human ace
11	63	42.9	145	7	ADC33366	Adc33366 Human nov
12	61.5	41.8	575	4	ABB11475	Abb11475 Human R31
13	61	41.5	98	3	AAB51865	Aab51865 Human sec
14	60	40.8	54	2	AAR37744	Aar37744 Collagen-
15	60	40.8	54	2	AAR93255	Aar93255 Collagen-
16	60	40.8	334	8	ADL91154	Adl91154 Mouse fic
17	60	40.8	334	8	ADL91152	Adl91152 Human fic
18	60	40.8	633	2	AAR37746	Aar37746 Collagen-
19	60	40.8	633	2	AAR93257	Aar93257 Collagen-
20	60	40.8	633	2	AAW57655	Aaw57655 Collagen-
21	60	40.8	1065	2	AAR37745	Aar37745 Collagen-
22	60	40.8	1065	2	AAR93256	Aar93256 Collagen-
23	60	40.8	1065	2	AAW57654	Aaw57654 Collagen-
24	60	40.8	1466	4	AAE02534	Aae02534 Bovine al
25	60	40.8	1466	4	AAE02533	Aae02533 Bovine al
26	59	40.1	532	2	AAW40114	Aaw40114 Human alp
27	58	39.5	226	7	ADB64162	Adb64162 Human pro
28	58	39.5	251	6	ADA54929	Ada54929 Human pro
29	58	39.5	393	5	AAU86144	Aau86144 Human PRO
30	58	39.5	393	6	ADA54950	Ada54950 Human pro
31	58	39.5	393	7	ADC31141	Adc31141 Human nov
32	58	39.5	393	7	ADJ37315	Adj37315 Human tum
33	58	39.5	393	8	ADG68239	Adg68239 Human PRO
34	58	39.5	421	7	ADF74163	Adf74163 Human nov
35	58	39.5	635	4	AAM78798	Aam78798 Human pro
36	58	39.5	638	5	ABJ01025	Abj01025 Human bre
37	58	39.5	638	6	ABU69146	Abu69146 Human NOV
38	58	39.5	638	8	ADO08295	Ado08295 Human NOV
39	58	39.5	644	4	AAM79782	Aam79782 Human pro
40	58	39.5	703	6	ABP96315	Abp96315 Human col
41	58	39.5	717	6	ABP96314	Abp96314 Human col
42	58	39.5	733	6	ABU69145	Abu69145 Human NOV
43	58	39.5	733	8	ADO08293	Ado08293 Human NOV
44	57.5	39.1	280	3	AAB19110	Aab19110 Polypepti
45	57	38.8	520	7	ABO80705	Abo80705 Pseudomon

ALIGNMENTS

RESULT 1
AAG62388
ID AAG62388 standard; peptide; 26 AA.
XX
AC AAG62388;
XX
DT 31-AUG-2001 (first entry)
XX
DE Alternatively splice AChE product C-terminus AChE-R.
XX
KW Antisense oligonucleotide; acetylcholine esterase; AChE; dystonia;
KW cholinergic neurotransmission; progressive neuromuscular disorder;
KW myasthenia gravis; Eaton-Lambert disease; muscular dystrophy; PTSD;
KW amyotrophic lateral sclerosis; post-traumatic stress disorder;
KW multiple sclerosis; post-stroke sclerosis; post-injury muscle damage;
KW excessive re-innervation.
XX
OS Unidentified.
XX
PN WO200136627-A2.
XX
PD 25-MAY-2001.
XX
PF 16-NOV-2000; 2000WO-IL000763.
XX
PR 16-NOV-1999; 99IL-00132972.
XX
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI Soreq H, Seidman S;
XX
DR WPI; 2001-336003/35.
XX
PT New antisense oligonucleotide targeted to acetylcholine esterase mRNA,
PT useful for treating or preventing progressive neuromuscular disorders
PT such as myasthenia gravis.
XX
PS Disclosure; Fig 1; 124pp; English.
XX
CC Sequences AAH44810 - AAH44822 represent antisense oligonucleotides
CC targeting the acetylcholine esterase (AChE) mRNA. AChE is involved in the
CC termination of cholinergic neurotransmission, by hydrolysing the
CC neurotransmitter acetylcholine. Mammalian AChE is encoded by one gene but
CC alternative splicing at its 3' end yields three different mRNA
CC transcripts which encode protein with distinct carboxyl termini. All
CC three proteins are catalytically active. AChE has morphogenic, non-
CC catalytic capacities too. AChE antisense oligonucleotides are used in
CC treating or preventing a progressive neuromuscular disorder. Examples of

CC disorders which are treatable using the antisense oligonucleotides
CC include myasthenia gravis, Eaton-Lambert disease, muscular dystrophy,
CC amyotrophic lateral sclerosis, post-traumatic stress disorder (PTSD),
CC multiple sclerosis, dystonia, post-stroke sclerosis, post-injury muscle
CC damage, excessive re-innervation and post-exposure to AChE inhibitors.
CC The present sequence represents the C-terminus of an alternatively
CC spliced AChE gene product
XX
SQ Sequence 26 AA;

Query Match 100.0%; Score 147; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.3e-10;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
||||||||||||||||||||||||||
Db 1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26

RESULT 2
AAB48914
ID AAB48914 standard; peptide; 26 AA.
XX
AC AAB48914;
XX
DT 16-MAR-2001 (first entry)
XX
DE Human acetylcholinesterase (AChE) C-terminal peptide, SEQ ID NO:1.
XX
KW Acetylcholinesterase; AChE; readthrough peptide; ARP; splice variant;
KW human; epitope; C-terminal peptide; antibody; central nervous system;
KW CNS stress; psychological insult; physical insult; chemical insult;
KW blood-brain barrier disruption; elevated glucocorticoid level;
KW Alzheimer's disease; diagnosis.
XX
OS Homo sapiens.
XX
PN WO200073343-A2.
XX
PD 07-DEC-2000.
XX
PF 31-MAY-2000; 2000WO-IL000312.
XX
PR 31-MAY-1999; 99IL-00130225.
XX
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI Soreq H, Kaufer D, Friedman A, Seidman S;
XX
DR WPI; 2001-061514/07.
XX
PT Antibody specific to acetylcholinesterase or its C-terminal peptide
PT derivative useful for diagnosing, ventral nervous system stress, elevated
PT glucocorticoid level, disruption of blood-brain barrier and Alzheimer's
PT disease.
XX
PS Claim 4; Page 42; 44pp; English.
XX
CC The invention relates to antibodies which recognise acetylcholinesterase
CC (AChE) or a C-terminal peptide thereof (particularly AAB48914-B48916).
CC The AChE splice variant, AChE-R, and AChE-R mRNA, have been found to be
CC elevated in response to central nervous system (CNS) insults. The
CC invention therefore also relates to a method for diagnosing CNS stress,
CC and also elevated glucocorticoid levels, disruption of the blood-brain
CC barrier or Alzheimer's disease using a sample (e.g., serum or
CC cerebrospinal fluid) and an antibody of the invention. The CNS stress
CC which may be diagnosed using the antibodies is preferably that caused by
CC psychological insult, physical insult (head injury, head trauma, or
CC exposure to irradiation) or chemical insult (exposure to insecticide or
CC nerve gas). The present sequence represents a human AChE C-terminal
CC peptide (termed AChE readthrough peptide (ARP) in the specification),
CC which is specifically claimed as an epitope which is recognised by an
CC antibody of the invention
XX
SQ Sequence 26 AA;

Query Match 100.0%; Score 147; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.3e-10;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
||||||||||||||||||||||||||
Db 1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26

RESULT 3
AAB50032
ID AAB50032 standard; peptide; 26 AA.
XX
AC AAB50032;
XX
DT 14-MAR-2001 (first entry)
XX
DE Acetylcholinesterase readthrough peptide ARP-1.
XX
KW ARP-1; haemostatic; acetylcholinesterase; AChE; cell growth; human;
KW cell differentiation; thrombocytopenia; post-irradiation condition;
KW post-chemotherapy condition; blood loss; stress-induced male infertility.
XX
OS Homo sapiens.
XX
PN WO200073427-A2.
XX
PD 07-DEC-2000.
XX
PF 31-MAY-2000; 2000WO-IL000311.
XX
PR 31-MAY-1999; 99IL-00130224.
PR 02-SEP-1999; 99IL-00131707.
XX
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI Soreq H, Eldor A, Deutch V, Grisaru D;
XX
DR WPI; 2001-061523/07.
XX
PT New regulatory peptides having cell growth and cell differentiation
PT activity derived from the C-terminal region of acetylcholinesterass
PT useful in promoting growth, survival and differentiation of stem cells.
XX
PS Claim 8; Page 50; 133pp; English.
XX
CC The present sequence is a C-terminal peptide of acetylcholinesterase
CC (AChE). This peptide is acetylcholinesterase "readthrough" peptide (ARP-
CC 1). This peptide has a cell growth and/or cell differentiation activity.
CC The peptide may be used in ex vivo or in vivo expansion of hsematopoietic
CC stem cells and neural progenitors, and in the promotion of megakaryocytic
CC differentiation of hematopoietic stem cells. In addition, the present
CC peptide may be used in for promoting expansion of committed neural
CC progenitors in a developing embryo, in cultured embryonic stem cells, and
CC embryoid bodies derived from them. The present peptide may further be
CC used in the treatment of thrombocytopenia, post-irradiation conditions,
CC post-chemotherapy conditions, and conditions following massive blood
CC loss, in inducing synthesis of AChE mRNA, and in promoting formation of
CC hematon bodies. Antibodies directed against the present peptide are
CC useful for diagnosing stress-induced male infertility
XX
SQ Sequence 26 AA;

Query Match 100.0%; Score 147; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.3e-10;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
||||||||||||||||||||||||||
Db 1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26

```
RESULT 4
AAU98024
ID   AAU98024 standard; peptide; 26 AA.
XX
AC   AAU98024;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human readthrough acetylcholinesterase unique region.
XX
KW   Human; acetylcholinesterase; single-chain variable fragment; scFv;
KW   AChE-S; heavy chain variable region; muscle re-innervation;
KW   progressive neuromuscular disorder; muscle distortion; readthrough;
KW   myasthenia gravis; neuromuscular junction abnormality;
KW   Eaton-Lambert disease; muscular dystrophy; amyotrophic lateral sclerosis;
KW   ALS; post-traumatic stress disorder; PTSD; multiple sclerosis; Dystonia;
KW   post-stroke sclerosis; post-injury muscle damage;
KW   excessive re-innervation.
XX
OS   Homo sapiens.
XX
PN   WO200246422-A1.
XX
PD   13-JUN-2002.
XX
PF   22-MAY-2001; 2001WO-IL000464.
XX
PR   04-DEC-2000; 2000IL-00140071.
XX
PA   (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI   Soreq H, Flores CF, Nissim A;
XX
DR   WPI; 2002-463832/49.
XX
PT   Nucleic acid sequence coding for a single-chain variable fragment (scFv)
PT   antibody that has specific affinity for the synaptic variant of
PT   acetylcholinesterase (AChE-S), useful for diagnosing a neuromuscular
PT   disorder, e.g. Myasthenia gravis.
XX
PS   Disclosure; Fig 1; 73pp; English.
XX
CC   The invention relates to a nucleic acid sequence coding for a single-
CC   chain variable fragment (scFv) antibody that has specific affinity for
CC   the synaptic variant of acetylcholinesterase (AChE-S), where the scFv
CC   antibody consists essentially of a polypeptide comprising the binding
CC   portion of the heavy chain variable region of an antibody. Also included
CC   are an expression vehicle comprising a nucleic acid sequence coding for a
CC   scFv antibody that has specific affinity for the synaptic variant of AChE
CC   -S, an scFv antibody specifically recognising and binding to the synaptic
CC   variant of AChE-S and a method for the diagnosis of a progressive
CC   neuromuscular disorder in a mammal, comprising obtaining a sample from
CC   the mammal and detecting intensified expression of at least one of the
CC   AChE variants in the sample. The single-chain Fv antibody is useful for
CC   diagnosing a progressive neuromuscular disorder which involves any one of
CC   muscle distortion, muscle re-innervation and neuromuscular junction (NMJ)
CC   abnormalities. The disorder is Myasthenia gravis (preferred), Eaton-
CC   Lambert disease, muscular dystrophy, amyotrophic lateral sclerosis (ALS),
CC   post-traumatic stress disorder (PTSD), multiple sclerosis, Dystonia, post
CC   -stroke sclerosis, post-injury muscle damage, excessive re-innervation,
CC   or post-exposure to AChE inhibitors. The present sequence represents the
CC   unique region of human acetylcholinesterase encoded by a readthrough
CC   splice variant
XX
SQ   Sequence 26 AA;

  Query Match             100.0%;  Score 147;  DB 5;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-10;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              ||||||||||||||||||||||||||
Db          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26


RESULT 5
AAW74588
ID   AAW74588 standard; protein; 31 AA.
XX
AC   AAW74588;
XX
DT   21-DEC-1998  (first entry)
XX
DE   Amino acid sequence of the human AChE variant 7.
XX
KW   Nuclease resistance; inhibition; human; acetyl-cholinesterase; AChE;
KW   central nervous system; CNS.
XX
OS   Homo sapiens.
XX
PN   WO9839486-A1.
XX
PD   11-SEP-1998.
XX
PF   06-MAR-1998;   98WO-US004503.
XX
PR   06-MAR-1997;   97US-0040203P.
XX
PA   (YISS ) YISSUM RES & DEV CO.
PA   (KOHN/) KOHN K I.
XX
PI   Soreq H,  Seidman S,  Shohami E;
XX
DR   WPI; 1998-506377/43.
XX
PT   Treatment of injury to central nervous system - by administration of
PT   inhibitor of acetyl-cholinesterase production.
XX
PS   Disclosure; Page 62; 88pp; English.
XX
CC   This is the amino acid sequence of a human acetyl-cholinesterase (AChE)
CC   variant used in the method of the invention, where inhibitors of AChE are
CC   used to treat injury to the central nervous system (CNS). The AChE
CC   inhibitor can also be used to facilitate transplantation of neuronal
CC   cells to the CNS of a patient. The inhibitor can also be used to improve
CC   hippocampal neuron survival following injury to the CNS. The CNS injuries
CC   that can be treated with the method include epilepsy, stroke,
CC   Huntington's disease, head injury, spinal injury, pain, Parkinson's
CC   disease, myelin deficiencies, neuromuscular disorders, neurological pain,
CC   amyotrophic lateral sclerosis, Alzheimer's disease, and affective
CC   disorders of the brain
XX
SQ   Sequence 31 AA;

  Query Match             100.0%;  Score 147;  DB 2;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-10;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              ||||||||||||||||||||||||||
Db          6 GMQGPAGSGWEEGSGSPPGVTPLFSP 31


RESULT 6
AAW68146
ID   AAW68146 standard; protein; 31 AA.
XX
AC   AAW68146;
XX
DT   05-OCT-1998  (first entry)
XX
DE   Human AChE splice variant E1-4-I4-E5.
XX
KW   Nuclease resistant; acetylcholinesterase; human; myasthenia gravis; AChE;
```

```
KW   Parkinson's disease; Alzheimer's disease; central nervous system;
KW   neuromuscular junction; cholinergic signalling; brain.
XX
OS   Homo sapiens.
XX
PN   WO9826062-A2.
XX
PD   18-JUN-1998.
XX
PF   12-DEC-1997;   97WO-US023598.
XX
PR   12-DEC-1996;   96US-0035266P.
PR   13-FEB-1997;   97US-0037777P.
PR   02-MAY-1997;   97US-00850347.
PR   21-JUL-1997;   97US-0053334P.
XX
PA   (YISS ) YISSUM RES & DEV CO.
PA   (KOHN/) KOHN K I.
XX
PI   Soreq H,  Seidman S,  Eckstein F,  Friedman A,  Kaufer D;
XX
DR   WPI; 1998-348522/30.
XX
PT   Synthetic nuclease resistant antisense oligodeoxynucleotides - directed
PT   against acetylcholinesterase, useful for treating Parkinson's and
PT   Alzheimer's diseases and myasthenia gravis.
XX
PS   Disclosure; Fig 12; 89pp; English.
XX
CC   This represents the amino acid sequence of a human acetylcholinesterase
CC   (AChE) splice variant. The invention provides sequences shown in AAV41278
CC   to AAV41285 that represent synthetic nuclease resistant antisense
CC   oligodeoxynucleotides which are capable of selectively modulating human
CC   acetylcholinesterase (AChE) production. These oligonucleotides are
CC   targeted to a splice junction in a splice variant of AChE mRNA and are
CC   capable of selectively modulating human AChE production in the central
CC   nervous system and neuromuscular junction. The invention also provides a
CC   method for determining the efficacy of these human AChE specific
CC   antisense oligonucleotides. These antisense oligonucleotides can be used
CC   to restore balanced cholinergic signalling in the brain, particularly
CC   related to learning and memory as well as stress disorders, Parkinson's
CC   and Alzheimer's disease. They can also be used to reduce production and
CC   therefore deposition of AChE in the neuromuscular junctions of patients
CC   with e.g. myasthenia gravis. The oligonucleotides work effectively at low
CC   doses while avoiding many of the side effects associated with Tacrine and
CC   related cholinergic drugs for Alzheimer's disease and pyridostigmine and
CC   related drugs for myasthenia gravis
XX
SQ   Sequence 31 AA;

  Query Match             100.0%;  Score 147;  DB 2;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-10;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              ||||||||||||||||||||||||||
Db          6 GMQGPAGSGWEEGSGSPPGVTPLFSP 31


RESULT 7
AAB50035
ID   AAB50035 standard; peptide; 37 AA.
XX
AC   AAB50035;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Acetylcholinesterase readthrough peptide ARP-2.
XX
KW   ARP-2; haemostatic; acetylcholinesterase; AChE; cell growth; human;
KW   cell differentiation; thrombocytopenia; post-irradiation condition;
KW   post-chemotherapy condition; blood loss; stress-induced male infertility.
XX
OS   Homo sapiens.
XX
PN   WO200073427-A2.
XX
PD   07-DEC-2000.
XX
PF   31-MAY-2000; 2000WO-IL000311.
XX
PR   31-MAY-1999;   99IL-00130224.
PR   02-SEP-1999;   99IL-00131707.
XX
PA   (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI   Soreq H,  Eldor A,  Deutch V,  Grisaru D;
XX
DR   WPI; 2001-061523/07.
XX
PT   New regulatory peptides having cell growth and cell differentiation
PT   activity derived from the C-terminal region of acetylcholinesterase
PT   useful in promoting growth, survival and differentiation of stem cells.
XX
PS   Example 10; Page 76; 133pp; English.
XX
CC   The present invention relates to C-terminal peptides of
CC   acetylcholinesterase (AChE) (see AAB50032-B50034). The peptides of the
CC   present invention have cell growth and/or cell differentiation activity.
CC   The peptides may be used in ex vivo or in vivo expansion of
CC   haematopoietic stem cells and neural progenitors, and in the promotion of
CC   megakaryocytic differentiation of hematopoietic stem cells. In addition,
CC   the peptides may be used in for promoting expansion of committed neural
CC   progenitors in a developing embryo, in cultured embryonic stem cells, and
CC   embryoid bodies derived from them. The peptides may further be used in
CC   the treatment of thrombocytopenia, post-irradiation conditions, post-
CC   chemotherapy conditions, and conditions following massive blood loss, in
CC   inducing synthesis of AChE mRNA, and in promoting formation of hematon
CC   bodies. Antibodies directed against the peptides are useful for
CC   diagnosing stress-induced male infertility. The present sequence is a C-
CC   terminal AChE "readthrough" peptide (ARP-2)
XX
SQ   Sequence 37 AA;

  Query Match             100.0%;  Score 147;  DB 4;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-10;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              ||||||||||||||||||||||||||
Db         12 GMQGPAGSGWEEGSGSPPGVTPLFSP 37


RESULT 8
AAB50036
ID   AAB50036 standard; protein; 53 AA.
XX
AC   AAB50036;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Acetylcholinesterase protein #1 used in a yeast two-hybrid system.
XX
KW   ARP; haemostatic; acetylcholinesterase; AChE; cell growth; human;
KW   cell differentiation; thrombocytopenia; post-irradiation condition;
KW   post-chemotherapy condition; blood loss; stress-induced male infertility.
XX
OS   Homo sapiens.
XX
PN   WO200073427-A2.
XX
PD   07-DEC-2000.
XX
PF   31-MAY-2000; 2000WO-IL000311.
XX
PR   31-MAY-1999;   99IL-00130224.
```

```
PR   02-SEP-1999;   99IL-00131707.
XX
PA   (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI   Soreq H,  Eldor A,  Deutch V,  Grisaru D;
XX
DR   WPI; 2001-061523/07.
XX
PT   New regulatory peptides having cell growth and cell differentiation
PT   activity derived from the C-terminal region of acetylcholinesterase
PT   useful in promoting growth, survival and differentiation of stem cells.
XX
PS   Claim 8; Page 87; 133pp; English.
XX
CC   The present invention relates to C-terminal peptides of
CC   acetylcholinesterase (AChE) (see AAB50032-B50034). The peptides of the
CC   present invention have cell growth and/or cell differentiation activity.
CC   The peptides may be used in ex vivo or in vivo expansion of
CC   haematopoietic stem cells and neural progenitors, and in the promotion of
CC   megakaryocytic differentiation of hematopoietic stem cells. In addition,
CC   the peptides may be used in for promoting expansion of committed neural
CC   progenitors in a developing embryo, in cultured embryonic stem cells, and
CC   embryoid bodies derived from them. The peptides may further be used in
CC   the treatment of thrombocytopenia, post-irradiation conditions, post-
CC   chemotherapy conditions, and conditions following massive blood loss, in
CC   inducing synthesis of AChE mRNA, and in promoting formation of hematon
CC   bodies. Antibodies directed against the peptides are useful for
CC   diagnosing stress-induced male infertility. The present sequence is a C-
CC   terminal AChE "readthrough" protein (ARP), which was used in a yeast two-
CC   hybrid system, to screen for ARP binding partners
XX
SQ   Sequence 53 AA;

  Query Match             100.0%;  Score 147;  DB 4;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-10;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              ||||||||||||||||||||||||||
Db         28 GMQGPAGSGWEEGSGSPPGVTPLFSP 53


RESULT 9
ABG31331
ID   ABG31331 standard; protein; 53 AA.
XX
AC   ABG31331;
XX
DT   05-NOV-2002  (first entry)
XX
DE   GFP-fused AChE variant expression construct, pGARP related protein.
XX
KW   Nervous system; drug assay; acetylcholinesterase; AChE; brain;
KW   isoform variance; AChE blocker; muscarinic receptor; M1; M2;
KW   pyridostigmine; muscarinic receptor blocker; scopolamine;
KW   M1 receptor blocker; pirenzepine; anxiety; post-traumatic stress;
KW   Alzheimer's disease; muscle malfunctioning; neurodegenerative disorder;
KW   xenobiotic damage; panic; neuromuscular disorder; Parkinson's disease;
KW   Huntington's chorea; muscle fatigue; multiple chemical sensitivity;
KW   autism; multiple sclerosis; Sjogren's disease; GFP; pGARP;
KW   green fluorescent protein.
XX
OS   Unidentified.
XX
PN   WO200240994-A2.
XX
PD   23-MAY-2002.
XX
PF   14-NOV-2001; 2001WO-IL001051.
XX
PR   14-NOV-2000; 2000US-0247970P.
XX
PA   (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
PI   Soreq H,  Meshorer E,  Sklan E,  Shoham S;
XX
DR   WPI; 2002-490152/52.
XX
PT   Evaluating effect of drugs on nervous system by comparing effect of drug
PT   on acetylcholinesterase, AChE activity in brain of test animal following
PT   challenge by AChE blocker and comparing it with control group.
XX
PS   Example; Page 52; 114pp; English.
XX
CC   The present invention relates to a method and system for evaluating an
CC   effect on the nervous system of a test drug. The method comprises
CC   comparing the effect of the drug on acetylcholinesterase (AChE) catalytic
CC   activity or isoform variance in a brain of a test animal following a
CC   challenge by an AChE blocker or a blocker of AChE and muscarinic
CC   receptors M1 and M2 (e.g. pyridostigmine) and comparing this effect with
CC   that of a known agent, preferably a non-selective muscarinic receptor
CC   blocker (e.g. scopolamine) or a specific M1 receptor blocker (e.g.
CC   pirenzepine). The method is useful for evaluating an effect on the
CC   nervous system of a test drug, including drugs for the treatment of
CC   anxiety conditions, post-traumatic stress, Alzheimer's disease, muscle
CC   malfunctioning, neurodegenerative disorders, damage resulting from
CC   exposure to xenobiotics, panic, neuromuscular disorders, Parkinson's
CC   disease, Huntington's chorea, muscle fatigue, multiple chemical
CC   sensitivity, autism, multiple sclerosis and Sjogren's disease. The
CC   present sequence represents a protein described in relation to green
CC   fluorescent protein (GFP)-fused AChE variant expression construct pGARP
CC   in the examples of the present invention
XX
SQ   Sequence 53 AA;

  Query Match             100.0%;  Score 147;  DB 5;  Length 53;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-10;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              ||||||||||||||||||||||||||
Db         28 GMQGPAGSGWEEGSGSPPGVTPLFSP 53


RESULT 10
AAW48797
ID   AAW48797 standard; protein; 600 AA.
XX
AC   AAW48797;
XX
DT   07-OCT-1998  (first entry)
XX
DE   Human acetylcholine esterase-I4 readthrough splice variant.
XX
KW   Human acetylcholine esterase-I4 readthrough splice variant; AChE-I4; CNS;
KW   blood/brain barrier; BBB; I4 peptide; antibiotic; brain tumour; glioma;
KW   chemotherapeutic drug; central nervous system.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .574
FT                   /note= "This region is encoded by exons 1-4 of AChE"
FT   Region          575. .599
FT                   /note= "I4 peptide encoded by intron 4 of AChE; this
FT                   sequence is claimed by the inventors under claim 2 in the
FT                   specification"
FT   Region          600
FT                   /note= "residue encoded by exon 5 of AChE"
XX
PN   WO9822132-A2.
XX
PD   28-MAY-1998.
XX
PF   20-NOV-1997;   97WO-US021696.
XX
```

PR 20-NOV-1996; 96US-0031194P.
PR 12-DEC-1996; 96US-0035266P.
PR 21-JUL-1997; 97US-0053200P.
XX
PA (YISS) YISSUM RES & DEV CO.
PA (KOHN/) KOHN K I.
XX
PI Soreq H, Friedman A, Seidmsn S, Kaufer D;
XX
DR WPI; 1998-312172/27.
XX
PT Increasing the permeability of the blood/brain barrier - using e.g.
PT adrenaline, stropine or acetylcholine esterase I4 splice vsriant peptide,
PT useful for imaging snd/or treatment of central nervous system disorders.
XX
PS Claim 1, 2; Page 45; 71pp; English.
XX
CC The present sequence represents human acetylcholine esterase-I4 (AChE-I4)
CC readthrough splice variant. The protein sequence comprises residues
CC encoded by exons 1-4 of human AChE followed by residues encoded by intron
CC 4, while the last residue of the protein is encoded by exon 5 of AChE.
CC The invention provides a pharmaceutical composition, for fscilitsting
CC passage of compounds through the blood/brain barrier (BBB), comprising of
CC AChE-I4, I4 peptide or AChE-I4 analogues together with a pharmaceutically
CC acceptable carrier. The pharmsceutical composition is claimed to
CC facilitate a reversible disruption of the BBB allowing transport of
CC compounds through the BBB. The compounds, e.g. imaging agents,
CC antibiotics or chemotherspeutic drugs, are claimed to be useful for the
CC disgnosis and treatment of disesses or disorders of the CNS such as
CC infections, neurochemical disorders, brain tumours, gliomas, etc
XX
SQ Sequence 600 AA;

Query Match 100.0%; Score 147; DB 2; Length 600;
Best Local Similarity 100.0%; Pred. No. 3.1e-09;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
||||||||||||||||||||||||||
Db 575 GMQGPAGSGWEEGSGSPPGVTPLFSP 600

RESULT 11
ADC33366
ID ADC33366 standard; protein; 145 AA.
XX
AC ADC33366;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human novel contig-encoded polypeptide sequence, SEQ ID NO:3448.
XX
KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnersry;
KW antiulcer; osteopathic; immunosuppressive; antiinflsmmatory; cytostatic;
KW gene therapy.
XX
OS Homo sapiens.
XX
PN WO2003029271-A2.
XX
PD 10-APR-2003.
XX
PF 24-SEP-2002; 2002WO-US030474.
XX
PR 24-SEP-2001; 2001US-0324631P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang TY, Zhsng J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX
DR WPI; 2003-371981/35.
DR N-PSDB; ADC32599.
XX
PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
PS Example 2; SEQ ID NO 3448; 1185pp; English.
XX
CC The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889)and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relstes to nucleic acid sequences over 99% identical with
CC ths novel human cDNAs. The invention additionslly encompasses expression
CC vectors and host cells comprising s nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; s method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying s compound which binds to s polypeptide of the invention. The
CC invention further discloses methods of peventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonsl antibodies for carrying out the methods of the
CC invention; methods for the identificstion of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disesse, Alzheimer's
CC disease and other neurodegenerative disesses, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune disesses or
CC cancer. The nucleic acids may also be used as hybridisstion probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequencs represents a human contig-
CC encoded polypeptide sequence used in an example of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specificstion, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 145 AA;

Query Match 42.9%; Score 63; DB 7; Length 145;
Best Local Similarity 54.2%; Pred. No. 8.3;
Matches 13; Conservative 3; Mismstches 8; Indels 0; Gaps 0;

Qy 3 QGPAGSGWEEGSGSPPGVTPLFSP 26
|| |: :|||||||| | : |
Db 71 QGDRGTEGKEGSGSPPGPTAGWMP 94

RESULT 12
ABB11475
ID ABB11475 standsrd; peptide; 575 AA.
XX
AC ABB11475;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human R31449_3 homologue, SEQ ID NO:1845.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulstion; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; ssthma; arthritis;

KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW sntifungal; vulnerary; antiulcer.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US003800.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-457740/49.
DR N-PSDB; ABA08719.
XX
PT Human proteins and DNA encoding sequences useful for preventing, treating
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
PT and cancer.
XX
PS Claim 20; Page 197; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention
XX
SQ Sequence 575 AA;

Query Match 41.8%; Score 61.5; DB 4; Length 575;
Best Local Similarity 56.5%; Pred. No. 50;
Matches 13; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

Qy 1 GMQG-PAGSGWEEGSGSPPGVTP 22
| || | | | |:| | |||: |
Db 421 GGQGLPEGWGLEKGEGLPPGIPP 443

RESULT 13
AAB51865
ID AAB51865 standard; protein; 98 AA.
XX
AC AAB51865;
XX
DT 16-FEB-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 39 SEQ ID NO:98.
XX
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW opthalmalogical; vulnerary; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorders; cancer; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
KW Alzheimer's disease; ocular disorder; wound healing; skin aging.
XX
OS Homo sapiens.
XX
PN WO200061626-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US009066.
XX
PR 09-APR-1999; 99US-0128698P.
PR 20-JAN-2000; 2000US-0176926P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-619227/59.
DR N-PSDB; AAC93517.
XX
PT New nucleic acid molecules encoding 49 human secreted proteins for
PT diagnosing, preventing or ameliorating medical conditions and used for
PT food additives or preservatives.
XX
PS Claim 11; Page 478; 516pp; English.
XX
CC Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding
CC human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -
CC AAB51927 represent alternative polypeptides encoded by the genes, and
CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include immunosuppressive;
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC virucide; fungicide; opthalmalogical; and vulnerary. The secreted
CC proteins, polynucleotides, antagonists and agonists may be useful in
CC treating, preventing and/or diagnosing diseases and disorders such as
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to

```
CC   increase or decrease storage capabilities, fat content, lipid, protein,
CC   carbohydrate, vitamins, minerals, cofactors and other nutritional
CC   components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are
CC   used in the isolation and characterisation of the proteins and
CC   polynucleotides of the invention
XX
SQ   Sequence 98 AA;

  Query Match             41.5%;  Score 61;  DB 3;  Length 98;
  Best Local Similarity   59.1%;  Pred. No. 9.7;
  Matches   13;  Conservative    0;  Mismatches    7;  Indels    2;  Gaps    1;

Qy          5 PAGSGWEEGSGSPPGVTPLFSP 26
              | |||||   | |  ||||  |
Db         64 PRGSGWERAPGCP--VTPLTLP 83


RESULT 14
AAR37744
ID   AAR37744 standard; protein; 54 AA.
XX
AC   AAR37744;
XX
DT   25-MAR-2003  (revised)
DT   07-SEP-1993  (first entry)
XX
DE   Collagen-like polymer DCP-(DB)3.
XX
KW   Recombinant; collagen-like polymer; CLP; tripeptide; helix; membrane;
KW   fibre; film; coating; triad sequence; collagen; mammalian; moulding;
KW   hydrogel; interchain linkage; colloid suspension; DCP; antibody.
XX
OS   Synthetic.
XX
PN   WO9310154-A1.
XX
PD   27-MAY-1993.
XX
PF   04-NOV-1992;   92WO-US009485.
XX
PR   12-NOV-1991;   91US-00791960.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J,  Ferrari FA;
XX
DR   WPI; 1993-182496/22.
DR   N-PSDB; AAQ43035.
XX
PT   High mol. wt. collagen-like protein polymers - capable of being produced
PT   in unicellular microorganisms.
XX
PS   Disclosure; Page 52; 82pp; English.
XX
CC   The sequences given in AAR37744 and AAR37747 represent examples of
CC   recombinantly produced DCP collagen-like polymers (CLPs) which consist of
CC   repeated tripeptide sequences selected from a wide range of GXY
CC   sequences, where X and Y can be any amino acid. The DNA encoding these
CC   sequences can be cloned into plasmids and used to transform E. coli to
CC   produce the DCP proteins. DCP peptides comprise repeated units of; B =
CC   GSRGDPGPP, C = GAHGPAGPK and/or D = GAQGPAGPG. These polymers have
CC   molecular weights of >30 kD and are able to form helices due to
CC   interchain linkages. These polymers pref. contain a proportion of
CC   tripeptide triad sequences found in natural collagens, pref. mammalian
CC   collagens. The CLPs impart unique characteristics to materials such as
CC   fibres, membranes, films, coatings, hydrogels, colloid suspensions and
CC   moulded articles. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 54 AA;

  Query Match             40.8%;  Score 60;  DB 2;  Length 54;
  Best Local Similarity   57.9%;  Pred. No. 7;
  Matches   11;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPG 19
              | |||||  |    |   |||
Db          1 GAQGPAGPGGSRGDPGPPG 19


RESULT 15
AAR93255
ID   AAR93255 standard; protein; 54 AA.
XX
AC   AAR93255;
XX
DT   25-MAR-2003  (revised)
DT   24-FEB-1997  (first entry)
XX
DE   Collagen-like polymer sequence D4/D5 unit (DB)3.
XX
KW   collagen; repetitive triad motif; recombinant production; photographic;
KW   medical; structural; fibre.
XX
OS   Synthetic.
XX
PN   US5496712-A.
XX
PD   05-MAR-1996.
XX
PF   05-NOV-1992;   92US-00972032.
XX
PR   06-NOV-1990;   90US-00609716.
PR   12-NOV-1991;   91US-00791960.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J,  Ferrari FA;
XX
DR   WPI; 1996-150728/15.
DR   N-PSDB; AAT16769.
XX
PT   Collagen-like polymers comprising repetitive triads - produced in
PT   unicellular organisms with improved characteristics, useful in, e.g.
PT   photographic and medical fibres.
XX
PS   Example 3; Col 23-24; 43pp; English.
XX
CC   The invention concerns collagen-like polymers having repetitive triads
CC   with reduced proline content, and where glycine is ths initial amino acid
CC   and the subsequent amino acids are varied. The choice of triads utilised
CC   in a recombinant collagen-like polymer are chosen in order to affect
CC   properties such as helix stability, hydration, solubility, gel point,
CC   biodegradation and immunogenicity. Triads of particular interest include
CC   GAP, GPA, GPP, GAS, GPG, GPS, GAQ, GSP, GLQ, GPR, GPK, GAK, GAR, GER,
CC   GDR, GEP, GDA, GAH and GEA. The collagen-like polymers may impart new
CC   characteristics, finding wide used in photographic, medical, structural
CC   and fibre applications, and are capable of being produced in unicellular
CC   microorganisms at high mol. wts. and in high efficiency. The present
CC   sequence, encoded by clone pPT0224 (see AAT16769), was identified to
CC   contain the SequenceD gene 4 or 5 monomer sequence (DB)3. The SequenceD
CC   collagen-like polymers are used as immunogens for the prepn. of
CC   antibodies. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 25-
CC   MAR-2003 to correct PA field.)
XX
SQ   Sequence 54 AA;

  Query Match             40.8%;  Score 60;  DB 2;  Length 54;
  Best Local Similarity   57.9%;  Pred. No. 7;
  Matches   11;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPG 19
              | |||||  |    |   |||
Db          1 GAQGPAGPGGSRGDPGPPG 19


Search completed: December 29, 2004, 14:58:43
```

Job time : 160 secs

THIS PAGE IS BLANK

OM protein - protein search, using sw model

Run on: December 29, 2004, 14:45:13 ; Search time 194 Seconds
(without alignments)
77.112 Million cell updates/sec

Title: US-09-998-042-1
Perfect score: 147
Sequence: 1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	67	45.6	1378	2	O97405	O97405 haliotis di
2	67	45.6	1449	2	Q910C0	Q910c0 oncorhynchu
3	62	42.2	234	2	Q6BP25	Q6bp25 debaryomyce
4	62	42.2	640	2	Q6KAQ4	Q6kaq4 mus musculu
5	62	42.2	640	2	BAD21403	Bad21403 mus muscu
6	62	42.2	699	2	Q6P1C4	Q6p1c4 mus musculu
7	62	42.2	699	2	AAH65148	Aah65148 mus muscu
8	62	42.2	1458	2	Q910B9	Q910b9 oncorhynchu
9	60	40.8	334	1	FCN1_MOUSE	O70165 mus musculu
10	60	40.8	1049	1	CA13_BOVIN	P04258 bos taurus
11	59.5	40.5	904	2	O76271	O76271 mytilus edu
12	59.5	40.5	905	2	Q8MW55	Q8mw55 mytilus gal
13	59	40.1	70	2	Q12985	Q12985 homo sapien
14	59	40.1	1447	2	Q6P4U1	Q6p4u1 brachydanio
15	59	40.1	1447	2	Q6U1J5	Q6u1j5 brachydanio
16	59	40.1	1447	2	AAH63249	Aah63249 brachydan
17	59	40.1	1447	2	AAR24536	Aar24536 brachydan
18	58.5	39.8	135	2	Q95LJ6	Q95lj6 macaca fasc
19	58.5	39.8	191	2	Q9CSM0	Q9csm0 mus musculu
20	58.5	39.8	386	2	Q01759	Q01759 pneumocysti
21	58.5	39.8	1076	2	Q01830	Q01830 pneumocysti
22	58	39.5	123	2	Q96ET3	Q96et3 homo sapien
23	58	39.5	164	2	Q7XDT8	Q7xdt8 oryza sativ
24	58	39.5	185	2	Q7XDT6	Q7xdt6 oryza sativ
25	58	39.5	185	2	Q7XDT7	Q7xdt7 oryza sativ
26	58	39.5	185	2	Q948R3	Q948r3 oryza sativ
27	58	39.5	251	2	Q96MW4	Q96mw4 homo sapien
28	58	39.5	313	2	Q96FI5	Q96fi5 homo sapien
29	58	39.5	322	2	Q9N5K5	Q9n5k5 caenorhabdi
30	58	39.5	393	2	Q7L8J4	Q7l8j4 homo sapien
31	58	39.5	393	2	AAQ88818	Aaq88818 homo sapi
32	58	39.5	421	2	Q9C0E3	Q9c0e3 homo sapien
33	58	39.5	591	2	O82178	O82178 arabidopsis
34	58	39.5	703	1	CA28_HUMAN	P25067 homo sapien
35	58	39.5	888	2	Q90796	Q90796 gallus gall
36	58	39.5	998	2	Q8CFM4	Q8cfm4 mus musculu
37	58	39.5	1222	2	Q8K173	Q8k173 mus musculu
38	58	39.5	1464	1	CA13_MOUSE	P08121 mus musculu
39	58	39.5	1464	2	Q7TT32	Q7tt32 mus musculu
40	58	39.5	1464	2	Q8BKY2	Q8bky2 mus musculu
41	58	39.5	1464	2	Q8BLW4	Q8blw4 mus musculu
42	57.5	39.1	196	2	Q25947	Q25947 plasmodium
43	57.5	39.1	280	2	Q8BP14	Q8bp14 mus musculu
44	57.5	39.1	280	2	Q8R138	Q8r138 mus musculu
45	57.5	39.1	2075	2	Q8JG44	Q8jg44 fugu rubrip

ALIGNMENTS

RESULT 1
O97405
ID O97405 PRELIMINARY; PRT; 1378 AA.
AC O97405;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Collagen pro alpha-chain precursor.
GN Name=Hdcol 1 alpha;
OS Haliotis discus (Abalone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Orthogastropoda;
OC Vetigastropoda; Haliotoidea; Haliotidae; Haliotis.
OX NCBI_TaxID=36094;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=99234051; PubMed=10215888;
RA Yoneda C., Hirayama Y., Nakaya M., Matsubara Y., Irie S., Hatae K.,
RA Watabe S.;
RT "The occurrence of two types of collagen proalpha-chain in the abalone
RT Haliotis discus muscle.";
RL Eur. J. Biochem. 261:714-721(1999).
DR EMBL; AB017600; BAA75668.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 17.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS50184; VWFC_2; 1.
KW Collagen; Signal.
FT SIGNAL 1 18 Potential.
SQ SEQUENCE 1378 AA; 132582 MW; 6EEE34A65B52E634 CRC64;

Query Match 45.6%; Score 67; DB 2; Length 1378;
Best Local Similarity 46.2%; Pred. No. 30;
Matches 12; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

```
Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              |: ||||     |   |||  |::||
Db       1111 GLTGPAGPSGPPGPPGPPGYGPVYSP 1136
```

RESULT 2
Q910C0
ID Q910C0 PRELIMINARY; PRT; 1449 AA.
AC Q910C0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

```
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Collagen a1(I).
GN   Name=COL1A1;
OS   Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21257802; PubMed=11358497;
RA   Saito M., Takenouchi Y., Kunisaki N., Kimura S.;
RT   "Complete primary structure of rainbow trout type I collagen
RT   consisting of a1(I)a2(I)a3(I) heterotrimers.";
RL   Eur. J. Biochem. 268:2817-2827(2001).
DR   EMBL; AB052835; BAB55661.1; -.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR   GO; GO:0006817; P:phosphate transport; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; VWC; 1.
DR   ProDom; PD000007; Clg_helix; 4.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Collagen.
SQ   SEQUENCE   1449 AA;  137116 MW;  62EEF8A7BFD652B8 CRC64;

  Query Match             45.6%;  Score 67;  DB 2;  Length 1449;
  Best Local Similarity   46.2%;  Pred. No. 31;
  Matches   12;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              |: ||:|   |:|:|.|||   |  |
Db        608 GVAGPSGERGEQGAGGPPGPQGLSGP 633


RESULT 3
Q6BP25
ID   Q6BP25      PRELIMINARY;      PRT;   234 AA.
AC   Q6BP25;
DT   01-OCT-2004 (TrEMBLrel. 28, Created)
DT   01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Similarity.
GN   ORFNames=DEHA0E17853g;
OS   Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX   NCBI_TaxID=4959;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS767;
RG   GENOLEVURES;
RA   Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA   Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA   Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA   Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA   Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA   Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA   Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA   Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA   Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA   Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA   Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA   Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA   Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA   Wincker P., Souciet J.L.;
RT   "Genome evolution in yeasts.";
RL   Nature 430:35-44(2004).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS767;
RA   Genoscope;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CR382137; CAG88303.1; -.
SQ   SEQUENCE   234 AA;  22636 MW;  3C0B851B5D0095C9 CRC64;

  Query Match             42.2%;  Score 62;  DB 2;  Length 234;
  Best Local Similarity   60.0%;  Pred. No. 19;
  Matches   12;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGV 20
              | ||| |:| | | | | ||
Db        126 GDQGPKGAGGERGIGGPQGV 145


RESULT 4
Q6KAQ4
ID   Q6KAQ4      PRELIMINARY;      PRT;   640 AA.
AC   Q6KAQ4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   MFLJ00201 protein (Fragment).
GN   Name=mFLJ00201;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinas; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryonic tail;
RA   Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA   Saga Y., Kitamura H., Nakagawa T., Nagase T., Ohara O., Koga H.;
RT   "Prediction of the Coding Sequences of Mouse Homologues of FLJ Genes:
RT   The Complete Nucleotide Sequences of 110 Mouse FLJ-Homologous cDNAs
RT   Identified by Screening of Terminal Sequences of cDNA Clones Randomly
RT   Sampled from Size-Fractionated Libraries.";
RL   DNA Res. 11:167-180(2004).
DR   EMBL; AK131153; BAD21403.1; -.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008983; TNF_like.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 7.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Collagen.
FT   NON_TER       1      1
SQ   SEQUENCE   640 AA;  61034 MW;  75CC9DEBAE5AC4B5 CRC64;

  Query Match             42.2%;  Score 62;  DB 2;  Length 640;
  Best Local Similarity   60.0%;  Pred. No. 52;
  Matches   12;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGV 20
              | :|| ||  | | | ||||
Db        307 GRRGPPGSKGEVGPGGPPGV 326


RESULT 5
BAD21403
ID   BAD21403    PRELIMINARY;      PRT;   640 AA.
AC   BAD21403;
DT   01-JUN-2004 (TrEMBLrel. 27, Created)
DT   01-JUN-2004 (TrEMBLrel. 27, Last sequencs update)
```

```
DT   01-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE   MFLJ00201 protein (Fragment).
GN   MFLJ00201.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryonic tail;
RA   Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA   Saga Y., Kitamura H., Nakagawa T., Nagase T., Ohara O., Koga H.;
RT   "Prediction of the Coding Sequences of Mouse Homologues of FLJ Gene:
RT   The Complete Nucleotide Sequences of 110 Mouse FLJ-homologous cDNAs
RT   Identified by Screening of Terminal sequences of cDNA Clones Randomly
RT   Sampled from Size-Fractionated Libraries.";
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK131153; BAD21403.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   640 AA;  61034 MW;  75CC9DEBAE5AC4B5 CRC64;

  Query Match             42.2%;  Score 62;  DB 2;  Length 640;
  Best Local Similarity   60.0%;  Pred. No. 52;
  Matches   12;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGV 20
              | :|| ||  | | | ||||
Db        307 GRRGPPGSKGEVGPGGPPGV 326


RESULT 6
Q6P1C4
ID   Q6P1C4      PRELIMINARY;      PRT;   699 AA.
AC   Q6P1C4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Procollagen, type VIII, alpha 2.
GN   Name=Col8a2;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC065148; AAH65148.1; -.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR008983; TNF_like.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 7.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Collagen.
SQ   SEQUENCE   699 AA;  66943 MW;  FCBD4FBB44642646 CRC64;

  Query Match             42.2%;  Score 62;  DB 2;  Length 699;
  Best Local Similarity   60.0%;  Pred. No. 57;
  Matches   12;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGV 20
              | :|| ||  | | | ||||
Db        366 GRRGPPGSKGEVGPGGPPGV 385


RESULT 7
AAH65148
ID   AAH65148    PRELIMINARY;      PRT;   699 AA.
AC   AAH65148;
DT   02-MAR-2004 (TrEMBLrel. 27, Created)
DT   02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT   02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE   Procollagen, type VIII, alpha 2.
GN   COL8A2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC065148; AAH65148.1; -.
KW   Collagen.
SQ   SEQUENCE   699 AA;  66943 MW;  FCBD4FBB44642646 CRC64;

  Query Match             42.2%;  Score 62;  DB 2;  Length 699;
  Best Local Similarity   60.0%;  Pred. No. 57;
  Matches   12;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGV 20
              | :|| ||  | | | ||||
Db        366 GRRGPPGSKGEVGPGGPPGV 385
```

```
RESULT 8
Q910B9
ID   Q910B9      PRELIMINARY;      PRT;  1458 AA.
AC   Q910B9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Collagen a3(I).
GN   Name=COL1A3;
OS   Oncorhynchus mykisa (Rainbow trout) (Salmo gairdneri).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchua.
OX   NCBI_TaxID=8022;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21257802; PubMed=11358497;
RA   Saito M., Takenouchi Y., Kunisaki N., Kimura S.;
RT   "Complete primary atructure of rainbow trout type I collagen
RT   conaisting of a1(I)a2(I)a3(I) heterotrimers.";
RL   Eur. J. Biochem. 268:2817-2827(2001).
DR   EMBL; AB052836; BAB55662.1; -.
DR   GO; GO:0005581; C:collagen; IEA.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0005201; F:extracellular matrix atructural constituent; IEA.
DR   GO; GO:0006817; P:phosphate transport; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   ProDom; PD000007; Clg_helix; 2.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Collagen.
SQ   SEQUENCE   1458 AA;  137757 MW;  AB1F9F3410A98650 CRC64;

  Query Match             42.2%;  Score 62;  DB 2;  Length 1458;
  Best Local Similarity   63.2%;  Pred. No. 1.2e+02;
  Matches   12;  Conaervative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPG 19
              | ||||||  |:| | | |
Db        611 GPQGPAGSQGEKGEGGPAG 629


RESULT 9
FCN1_MOUSE
ID   FCN1_MOUSE     STANDARD;      PRT;   334 AA.
AC   O70165;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Laat sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Ficolin 1 precursor (Collagen/fibrinogen domain-containing protein 1)
DE   (Ficolin-A) (Ficolin A) (M-Ficolin).
GN   Name=Fcn1; Synonyms=Fcna;
OS   Mus muaculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=98205801; PubMed=9535745;
RA   Fujimori Y., Harumiya S., Fukumoto Y., Miura Y., Yagasaki K.,
RA   Tachikawa H., Fujimoto D.;
RT   "Molecular cloning and characterization of mouse ficolin-A.";
RL   Biochem. Biophys. Res. Commun. 244:796-800(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collina F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richarda S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibba R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodriguea S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakealey R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinaki M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analyais of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Involved in serum exerting lectin activity. Binds GlcNAc
CC       (By similarity).
CC   -!- SUBUNIT: Homopolymer. Interacts with elaatin (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted. Found on the monocyte surface (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: Highly expreased in liver and apleen.
CC   -!- SIMILARITY: Belongs to the ficolin lectin family.
CC   -!- SIMILARITY: Contains 1 collagenous domain.
CC   -!- SIMILARITY: Containa 1 fibrinogen C-terminal domain.
CC   ----------------------------------------------------------------------
CC   Thia SWISS-PROT entry is copyright. It ia produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Inatitute.  There are no  reatrictions on  its
CC   uae  by  non-profit  inatitutiona as long  aa its content  is  in  no  way
CC   modified and this atatement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-aib.ch/announce/
CC   or send an email to licenae@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB007813; BAA25126.1; -.
DR   EMBL; BC019180; AAH19180.1; -.
DR   PIR; JC5980; JC5980.
DR   HSSP; P02671; 1FZD.
DR   MGD; MGI:1340905; Fcna.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   Pfam; PF01391; Collagen; 1.
DR   Pfam; PF00147; Fibrinogen_C; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   SMART; SM00186; FBG; 1.
DR   PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; 1.
KW   Collagen; Glycoprotein; Lectin; Multigene family; Repeat; Signal.
FT   SIGNAL       1     22       Potential.
FT   CHAIN       23    334       Ficolin 1.
FT   DOMAIN      50     88       Collagen-like.
FT   DOMAIN     152    298       Fibrinogen C-terminal.
FT   CARBOHYD   261    261       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   334 AA;  36298 MW;  9D30C05036AA04B1 CRC64;

  Query Match             40.8%;  Score 60;  DB 1;  Length 334;
  Best Local Similarity   60.0%;  Pred. No. 47;
  Matchea   12;  Conservative    1;  Mismatches    7;  Indela    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGV 20
              |  |||||  | |:  ||||
Db         77 GKMGPAGSKGEPGTMGPPGV 96


RESULT 10
```

```
CA13_BOVIN
ID   CA13_BOVIN     STANDARD;      PRT;  1049 AA.
AC   P04258;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Collagen alpha 1(III) chain.
GN   Name=COL3A1;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 1-242.
RX   MEDLINE=80026026; PubMed=488906;
RA   Fietzek P.P., Allmann H., Rauterberg J., Henkel W., Wachter E.,
RA   Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. I. The amino
RT   acid sequence of the amino terminal region of the alpha 1(III) chain
RT   (positions 1-222).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).
RN   [2]
RP   SEQUENCE OF 243-422.
RX   MEDLINE=80026027; PubMed=488907;
RA   Dewes H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. II. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2
RT   (positions 223-402).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).
RN   [3]
RP   SEQUENCE OF 423-571.
RX   MEDLINE=80026028; PubMed=488908;
RA   Bentz H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. III. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
RT   (positions 403-551).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).
RN   [4]
RP   SEQUENCE OF 572-808.
RX   MEDLINE=80026029; PubMed=488909;
RA   Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. IV. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
RT   (positions 552-788).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).
RN   [5]
RP   SEQUENCE OF 809-947.
RX   MEDLINE=80026030; PubMed=488910;
RA   Dewes H., Fietzek P.P., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. V. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
RT   (position 789-927).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).
RN   [6]
RP   SEQUENCE OF 948-1049.
RX   MEDLINE=80026031; PubMed=488911;
RA   Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;
RT   "The covalent structure of calf skin type III collagen. VI. The amino
RT   acid sequence of the carboxyterminal cyanogen bromide peptide alpha
RT   1(III)CB9B (positions 928-1028).";
RL   Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).
CC   -!- FUNCTION: Collagen type III occurs in most soft connective tissues
CC       along with type I collagen.
CC   -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC       linked to each other by interchain disulfide bonds. Trimers are
CC       also cross-linked via hydroxylysines.
CC   -!- PTM: Prolines at the third position of the tripeptide repeating
CC       unit (G-X-Y) are hydroxylated in some or all of the chains.
CC   -!- SIMILARITY: Belongs to the fibrillar collagen family.
DR   PIR; A02862; CGBO7S.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01391; Collagen; 17.
DR   ProDom; PD000007; Clg_helix; 3.
DR   PROSITE; PS01208; VWFC_1; PARTIAL.
KW   Collagen; Connective tissue; Direct protein sequencing;
KW   Extracellular matrix; Glycoprotein; Hydroxylation; Repeat.
FT   DOMAIN        1     14       Nonhelical region (N-terminal).
FT   DOMAIN       15   1040       Triple-helical region.
FT   DOMAIN     1041   1049       Nonhelical region (C-terminal).
FT   MOD_RES      95     95       5-hydroxylysine.
FT   MOD_RES     107    107       5-hydroxylysine.
FT   MOD_RES     119    119       5-hydroxylysine.
FT   MOD_RES     938    938       5-hydroxylysine.
FT   MOD_RES     950    950       5-hydroxylysine.
FT   CARBOHYD    107    107       O-linked (Gal. . .).
FT   CARBOHYD    950    950       O-linked (Gal. . .).
FT   DISULFID   1040   1040       Interchain.
FT   DISULFID   1041   1041       Interchain.
SQ   SEQUENCE   1049 AA;  93651 MW;  8EEC33D1C66EC9A3 CRC64;

  Query Match             40.8%;  Score 60;  DB 1;  Length 1049;
  Best Local Similarity   57.1%;  Pred. No. 1.5e+02;
  Matches   12;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGVT 21
              | |||||   | |   ||| |
Db        447 GPQGPAGKNGETGPQGPPGPT 467


RESULT 11
O76271
ID   O76271      PRELIMINARY;      PRT;   904 AA.
AC   O76271;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nongradient byssal.
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
OX   NCBI_TaxID=6550;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98393676; PubMed=9724735;
RA   Qin X.X., Waite J.H.;
RT   "A potential mediator of collagenous block copolymer gradients in
RT   mussel byssal threads.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:10517-10522(1998).
DR   EMBL; AF043944; AAC33847.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0006817; P:phosphate transport; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   Pfam; PF01391; Collagen; 7.
DR   ProDom; PD000007; Clg_helix; 1.
KW   Collagen.
SQ   SEQUENCE   904 AA;  77883 MW;  5529135651AD4C40 CRC64;

  Query Match             40.5%;  Score 59.5;  DB 2;  Length 904;
  Best Local Similarity   50.0%;  Pred. No. 1.5e+02;
  Matches   13;  Conservative    3;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              | :||||   ::| | |||  |  ||
Db        227 GPRGPAGPPGDQGHGGPPG-PPGHSP 251


RESULT 12
Q8MW55
ID   Q8MW55      PRELIMINARY;      PRT;   905 AA.
AC   Q8MW55;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
```

```
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Precollagen-NG.
OS   Mytilus galloprovincialis (Mediterranean mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
OX   NCBI_TaxID=29158;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22038007; PubMed=12042339;
RA   Lucas J.M., Vaccaro E., Waite J.H.;
RT   "A molecular, morphometric and mechanical comparison of the structural
RT   elements of byssus from Mytilus edulis and Mytilus
RT   galloprovincialis.";
RL   J. Exp. Biol. 205:1807-1817(2002).
DR   EMBL; AF448524; AAM34599.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0006817; P:phosphate transport; IEA.
DR   InterPro; IPR008160; Collagen.
DR   Pfam; PF01391; Collagen; 7.
KW   Collagen.
SQ   SEQUENCE   905 AA;  79251 MW;  499F6F6E26A5C182 CRC64;

  Query Match             40.5%;  Score 59.5;  DB 2;  Length 905;
  Best Local Similarity   50.0%;  Pred. No. 1.5e+02;
  Matches   13;  Conservative    3;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              | :||||    ::| | |||  |  ||
Db        232 GPRGPAGPPGDQGHGGPPG-PPGHSP 256


RESULT 13
Q12985
ID   Q12985      PRELIMINARY;      PRT;    70 AA.
AC   Q12985;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Alpha-1 type II collagen (Fragment).
GN   Name=COL2A1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=95150028; PubMed=7847372;
RA   Tiller G.E., Weis M.A., Polumbo P.A., Gruber H.E., Rimoin D.L.,
RA   Cohn D.H., Eyre D.R.;
RT   "An RNA-splicing mutation (G+5IVS20) in the type II collagen gene
RT   (COL2A1) in a family with spondyloepiphyseal dysplasia congenita.";
RL   Am. J. Hum. Genet. 56:388-395(1995).
DR   EMBL; U15195; AAB60370.1; -.
DR   PIR; A38513; CGHU6C.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0006817; P:phosphate transport; IEA.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   Pfam; PF01391; Collagen; 1.
DR   ProDom; PD000007; Clg_helix; 1.
KW   Collagen.
FT   NON_TER       1      1
FT   NON_TER      70     70
SQ   SEQUENCE   70 AA;  6482 MW;  13AE3044C3F2FC07 CRC64;

  Query Match             40.1%;  Score 59;  DB 2;  Length 70;
  Best Local Similarity   46.7%;  Pred. No. 12;
  Matches   14;  Conservative    1;  Mismatches   11;  Indels    4;  Gaps    1;

Qy          1 GMQGPAGSGWEEGS----GSPPGVTPLFSP 26
              | |||  |   |||     | | ||  :  |
Db         38 GPQGPPGPAGEEGKRGARGEPGGVGPIGPP 67


RESULT 14
Q6P4U1
ID   Q6P4U1      PRELIMINARY;      PRT;  1447 AA.
AC   Q6P4U1;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Col1a1 protein.
GN   Name=col1a1;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Strausberg R.;
RL   Submitted (DEC-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC063249; AAH63249.1; -.
DR   InterPro; IPR008161; Clg_helix.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR009041; PMP_SGCI.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF00093; VWC; 1.
DR   ProDom; PD000007; Clg_helix; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Collagen.
SQ   SEQUENCE   1447 AA;  136955 MW;  74723FC8ACAAAD86 CRC64;

  Query Match             40.1%;  Score 59;  DB 2;  Length 1447;
  Best Local Similarity   46.2%;  Pred. No. 2.7e+02;
  Matches   12;  Conservative    2;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              |  ||||   |:|:  |||   |  |
Db        607 GPAGPAGERGEQGAAGPPGFQGLPGP 632
```

```
RESULT 15
Q6U1J5
ID   Q6U1J5      PRELIMINARY;      PRT;  1447 AA.
AC   Q6U1J5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Chihuahua.
GN   Name=chi;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22984997; PubMed=14623232;
RA   Fisher S., Jagadeeswaran P., Halpern M.E.;
RT   "Radiographic analysis of zebrafish skeletal defects.";
RL   Dev. Biol. 264:64-76(2003).
DR   EMBL; AY380817; AAR24536.1; -.
DR   InterPro; IPR008160; Collagen.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR009041; PMP_SGCI.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01410; COLFI; 1.
DR   Pfam; PF01391; Collagen; 17.
DR   Pfam; PF00093; VWC; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Collagen.
SQ   SEQUENCE   1447 AA;  137144 MW;  9CABD561F5BA36BF CRC64;

  Query Match             40.1%;  Score 59;  DB 2;  Length 1447;
  Best Local Similarity   46.2%;  Pred. No. 2.7e+02;
  Matches   12;  Conservative    2;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 GMQGPAGSGWEEGSGSPPGVTPLFSP 26
              |  ||||   |:|:  |||   |  |
Db        607 GPAGPAGERGEQGAAGPPGFQGLPGP 632


Search completed: December 29, 2004, 15:02:08
Job time : 201 secs
```

THIS PAGE IS BLANK